Multinomial Processing Tree (MPT) Modeling: Basics

Daniel W. Heck

Philipps-Universität Marburg

(slides adapted from Edgar Erdfelder)

1) Basics

- 1.1) Introduction to MPT models
- 1.2) Examples
- 1.3) Model development
- 1.4) Formal model structure
- 1.5) Identifiability
- 1.6) Parameter estimation
- 1.7) Model assessment

1.1) Introduction to MPT Models

Required type of data:

- Multinomial models are tailored to discrete (i.e., categorical) data.
  - yes/no responses, correct/incorrect judgments, ratings, choices, ...
- Psychological data are typically discrete in nature
  - If not, they can be transformed into discrete data
  - Response time bins, rankings of numerical judgments, …
- Hence, many psychological paradigms generate frequency data that are appropriate for MPT modeling.

Measurement of Cognitive Processes

MPT models…

- … provide **explanations** of observed frequency data in terms of basic parameters with clear-cut psychological interpretations;
- … these **parameters** represent probabilities of latent psychological processes (or latent psychological states) underlying human behavior;
- … in other words, these models **disentangle and measure** the contributions of different psychological processes to frequencies of observable behaviors.

1.1) Introduction to MPT models

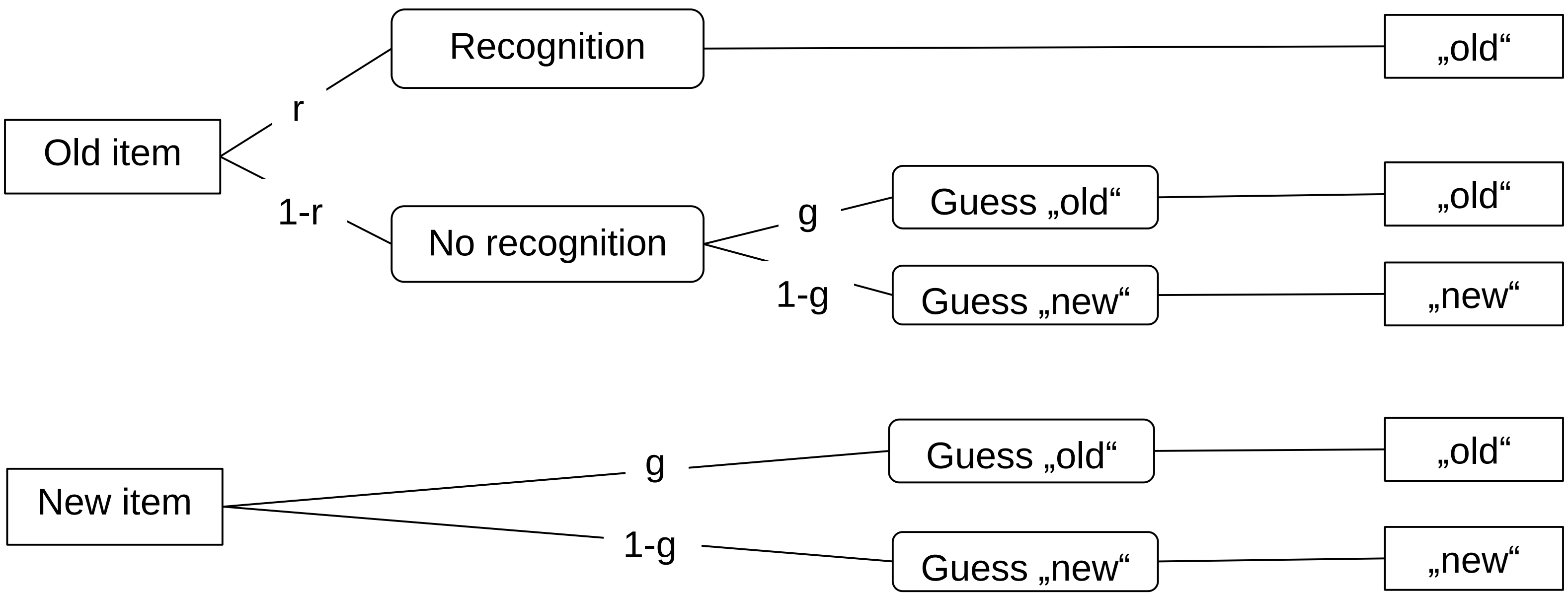
- **“Multinomial”**
  - MPT models assume that observations are sampled independently from one or more multinomial distributions
  - The frequency data structure can be univariate or multivariate
- “**Processing**”
  - Assumption that a finite number of latent processes generate the observed responses
  - Goal: Drawing inferences about these processes (e.g., via parameter estimation or hypothesis testing)
- “**Tree**”
  - Models can be depicted as probability trees

1.2) Examples

A very simple example:

- *Paradigm:*
  - Yes-No recognition test
- *Two Conditions:*
  - Old Items
  - New Items
- *Categorical (dichotomous) dependent variable:*
  - „old“ vs. „new“ Judgment

A) One-High Threshold Model (Blackwell, 1963)

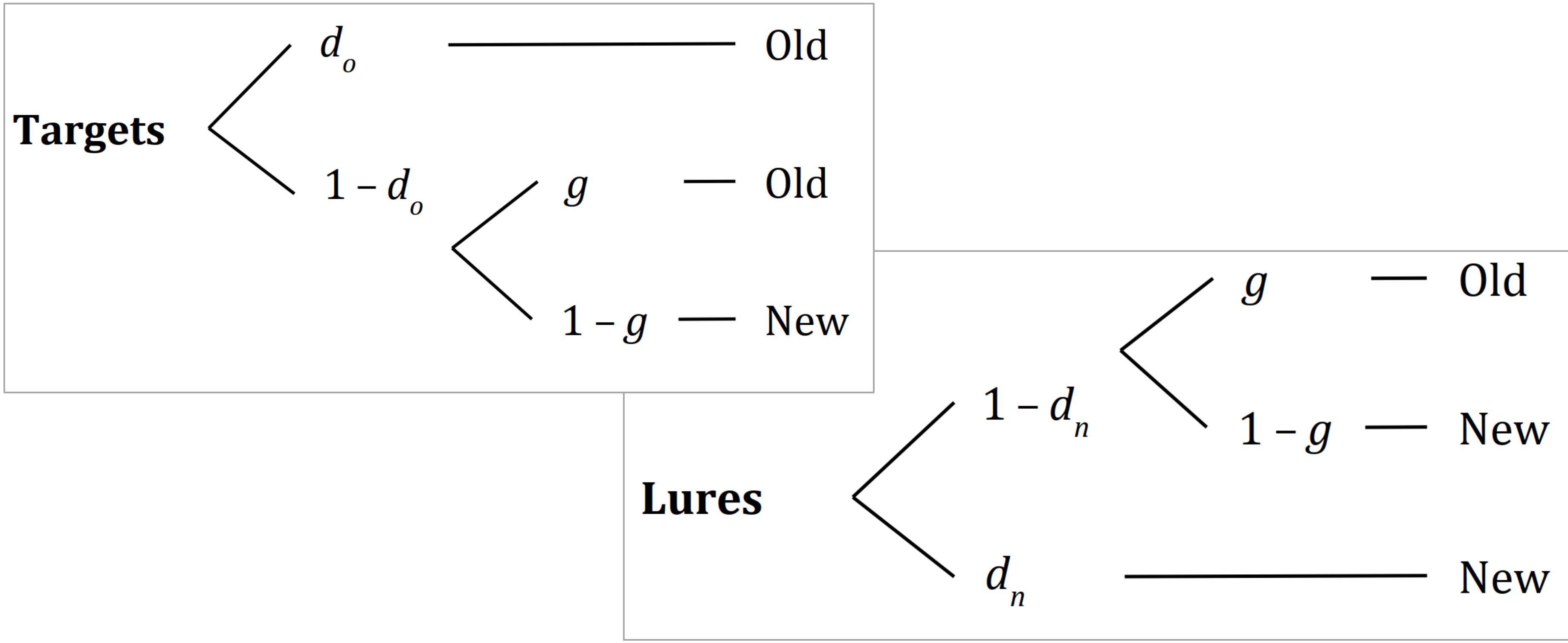


Model equations:

p(„old“ | old item) = $r + (1\text{-}r) \cdot g$

p(„old“ | new item) = g

B) Two-High Threshold Model



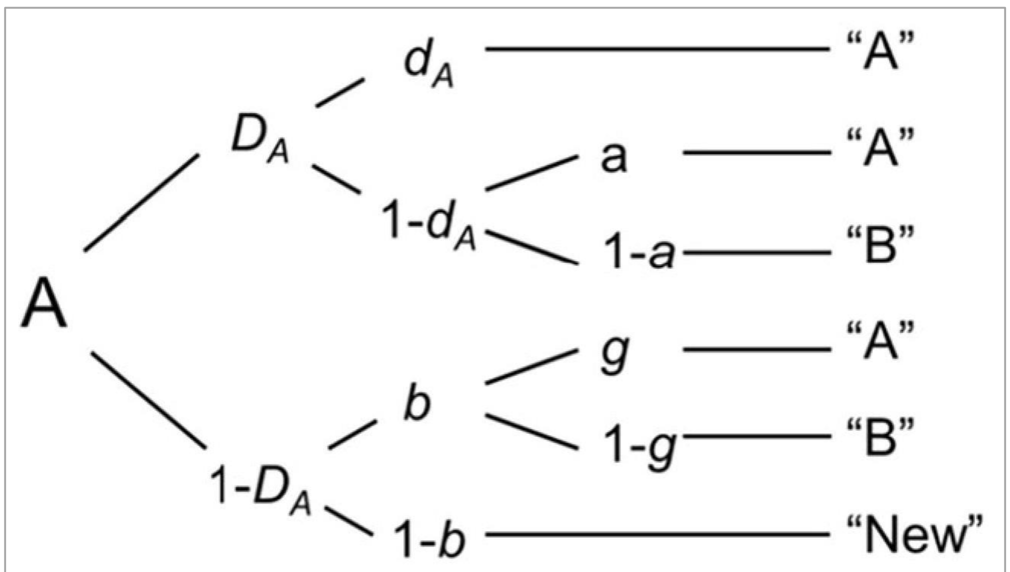
Model equations:

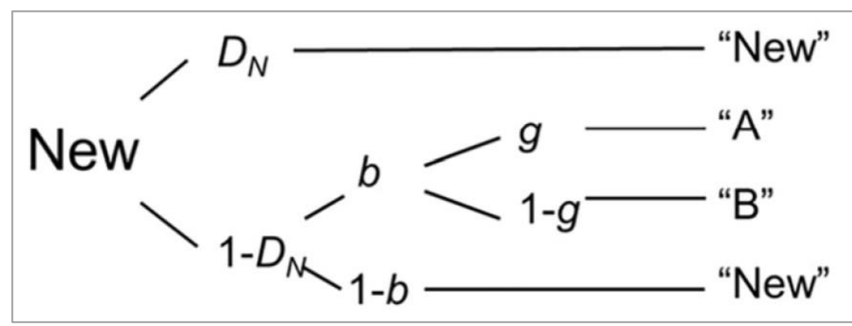
p(„old“ | old item) = $d_o + (1\text{-}d_o) \cdot g$

p(„old“ | new item) = $d_n + (1\text{-}d_n) \cdot (1\text{-}g)$

C) Source-Monitoring Model

- *Paradigm:* Recognition test with the two Sources A and B
- *Conditions*: Test items from Source A or B, and New items
- *Categorical dependent variable*: Participants' responses whether a presented item is from Source "A", "B", or "New"





1.3) Model Development: Part I

- Necessary steps:
  1. Select a paradigm (e.g., a task)
  2. Define the conditions of the paradigm
  3. Define the category system for each condition
  4. List relevant processes/parameters
  5. Construct theoretically reasonable processing branches („trees“) for each condition
  6. Derive corresponding model equations.

- General rules:
  - As simple as possible!!
  - Ignore unlikely events

1.3) Model Development: Part II

- Technical issues:
  - Identifiability: Is it possible to obtain unique parameter estimates?
  - Statistical power: How many observations are required?

- Substantive issue: Construct validity
  - Empirical validation of MPT model parameters via selective influence:
  - Experimental manipulation should selectively influence one specific parameter but no other parameters

1.4) Formal Model Structure

Simple multinomial model:

- One variable with J categories
- Observed frequencies: $n_1, n_2, ..., n_J$
- Vector of category probabilities: $\boldsymbol{p} = (p_1, p_2, ..., p_J)$
- Given independent sampling, the frequencies follow a multinomial distribution:

$$p_{N,\pi}(n_1, n_2, \ldots, n_J) = \frac{N}{n_1! n_2! \ldots n_J!} p_1^{n_1} p_2^{n_2} \ldots p_J^{n_J}$$

1.4) Parameterized Multinomial Models

- The category probabilities p_1, p_2 etc. are rewritten as functions of "**latent parameters**" $\theta_1, \theta_2, ..., \theta_S$

- Based on the simple multinomial model, we define a set of "**model equations**":
  - $p_1 = f_1(\theta_1, \theta_2, ..., \theta_S)$
  - $p_2 = f_2(\theta_1, \theta_2, ..., \theta_S)$
  - ....
  - $p_J = f_J(\theta_1, \theta_2, ..., \theta_S)$

- The set of possible values of S latent parameters is called "**parameter space**" Ω of the model.

1.4) Parameterized Multinomial Models

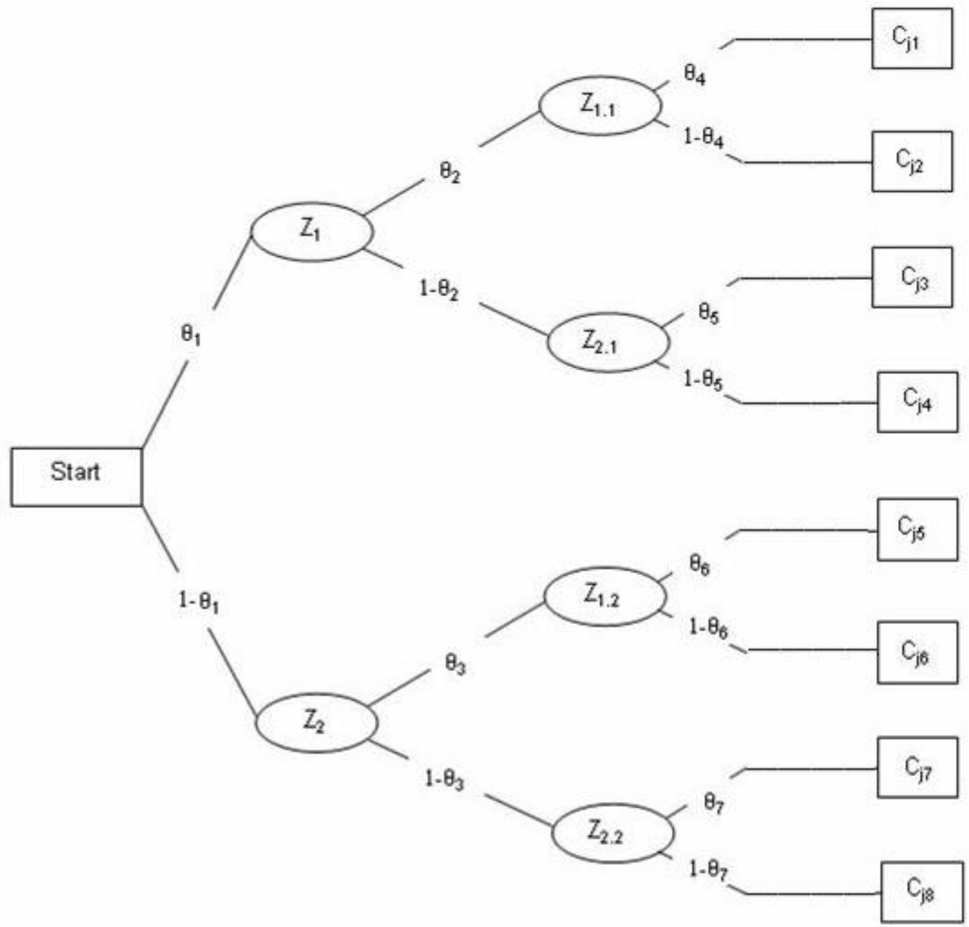
- *Example*: Model equations of the 1HTM

```
target: p(hit)    = r + (1−r) g
target: p(miss)   = (1−r) (1−g)

lure:   p(false alarm)         = g
lure:   p(correct rejection)   = (1−g)
```

$$P(n_{hit}, n_{miss} \mid r, g) = \frac{N}{n_{hit}!\, n_{miss}!} \boxed{[r + (1 - r) g]}^{n_{hit}} \boxed{[(1 - \mathrm{r})\,(1 - \mathrm{g})]}^{n_{miss}}$$

MPT models assume a *specific form* of the model equations (i.e., a binary probability tree)



1.4) Formal Definition of MPT Models

- Formally, any MPT model can be represented as:

$$p_j = \sum_{i=1}^{I(j)} c_{ij} \prod_{s=1}^{S} \theta_s^{a_{ijs}} \cdot (1-\theta_s)^{b_{ijs}}, \qquad \sum_{j=1}^{J} p_j = 1, \qquad \theta_s \in [0,\ 1]$$

where s : Parameter index

j : Category index

i : Branch index

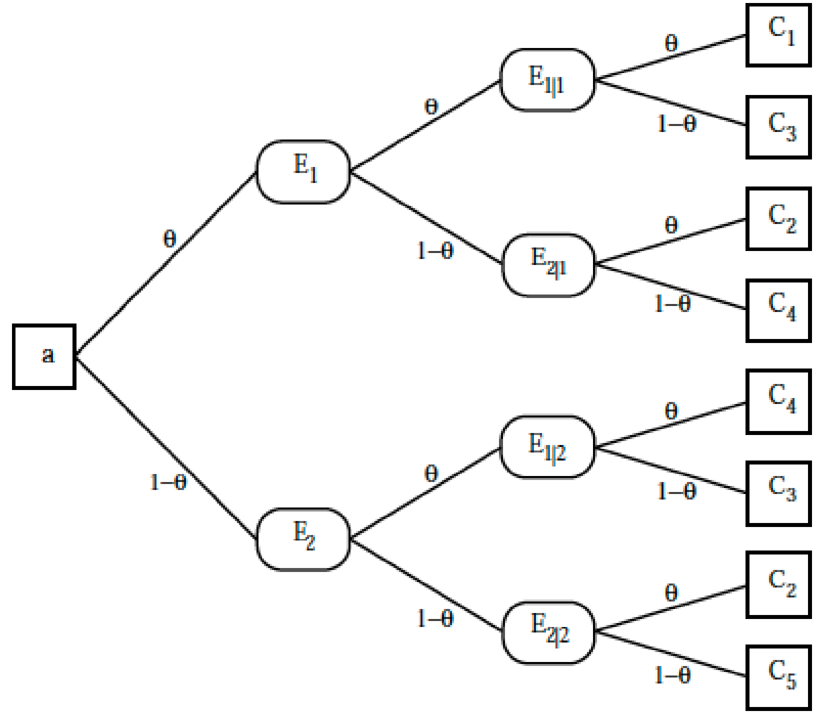
c_{ij} : positive real number

a_{ijs}, b_{ijs} : nonnegative integer number (often 0 or 1)

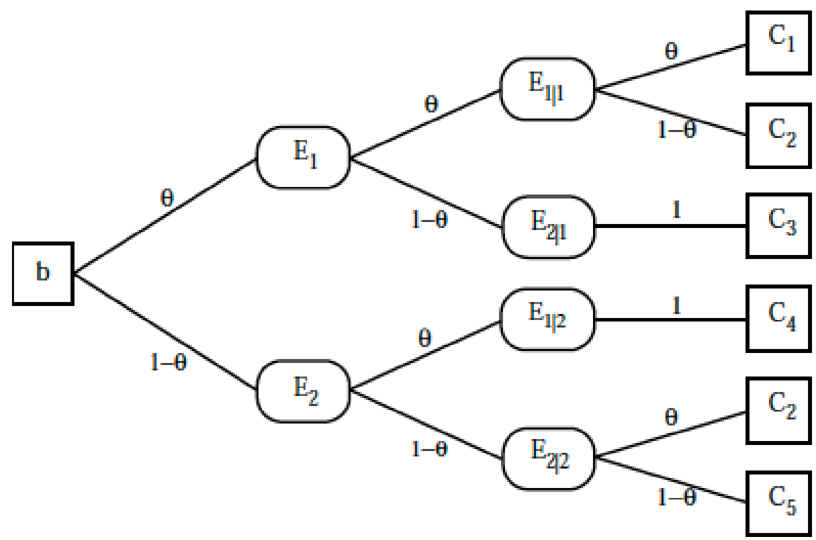
Uniqueness of the „Tree“

- A binary probabilistic event tree uniquely determines a system of MPT model equations.
- However: it is not true that any system of MPT model equations uniquely determines a specific processing tree diagram.
- Counter Example 1: Level switching in independence models

Counter Example 2



$$p_1(\theta) = \theta^3$$
$$p_2(\theta) = \theta \cdot (1 - \theta)$$
$$p_3(\theta) = \theta \cdot (1 - \theta)$$
$$p_4(\theta) = \theta \cdot (1 - \theta)$$
$$p_5(\theta) = (1 - \theta)^3$$



1.5) Identifiability

- MPT models define a mapping $f: \Omega \rightarrow P$

Ω is called „**Parameter Space**“:

= Set of all possible parameter vectors

P is called „**Data Space**“ (more precisely: space of category probabilities)

= Set of all possible category probability vectors

1.5) Identifiability

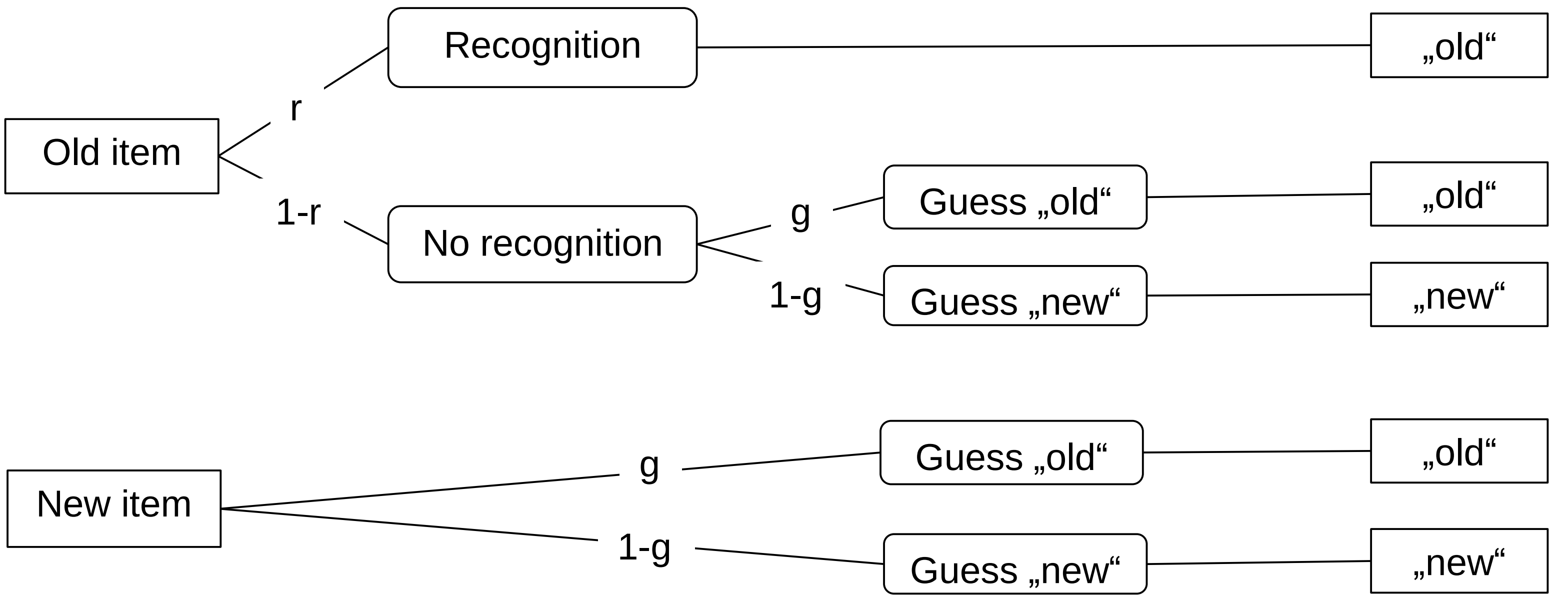
Definition (**global identifiability**):

A MPT model is globally identified if *f* is one-to-one.

Definition (**local identifiability**):

A MPT model is locally identified if *f* is one-to-one in the neighborhood of θ_0 in Ω.

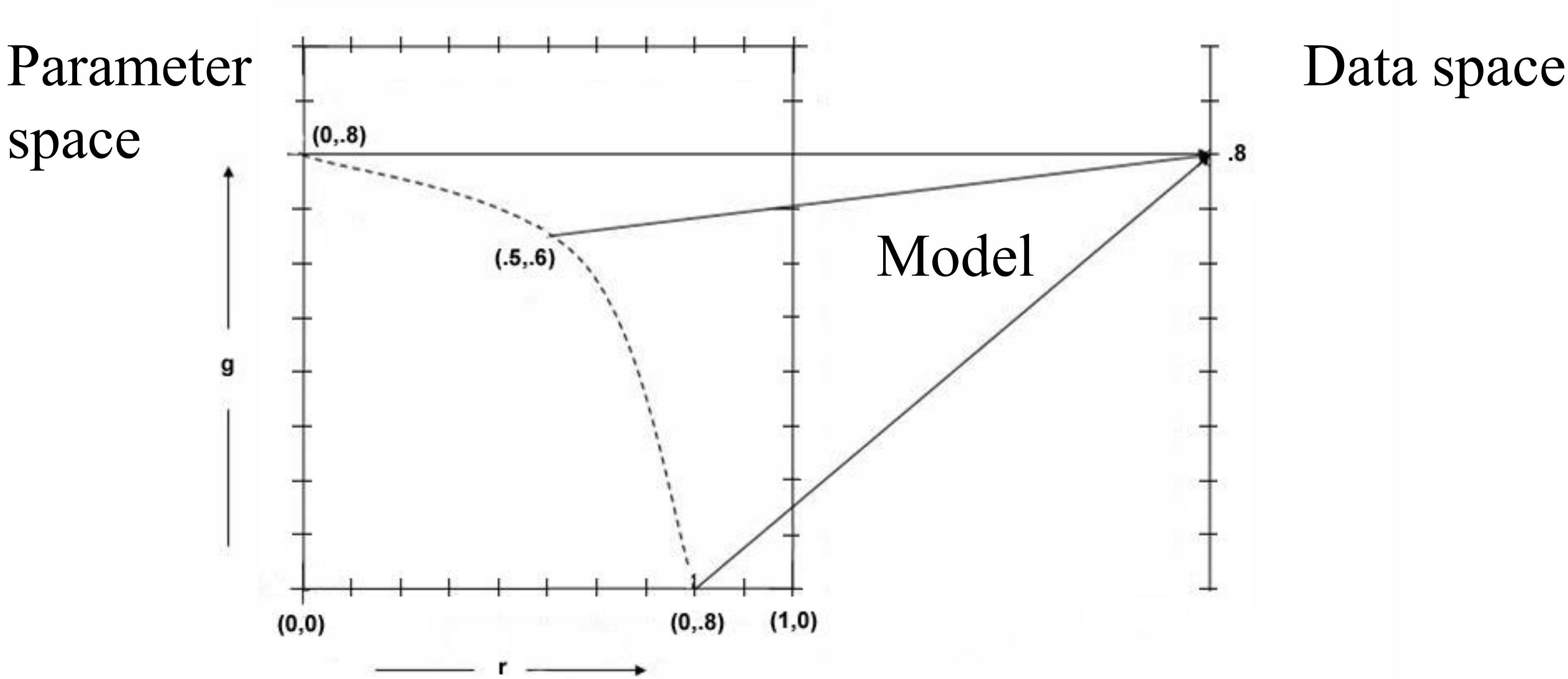
One-High-Threshold Recognition Model (Blackwell, 1953)



r: probability of recognition

g: probability of guessing „old“ given no recognition

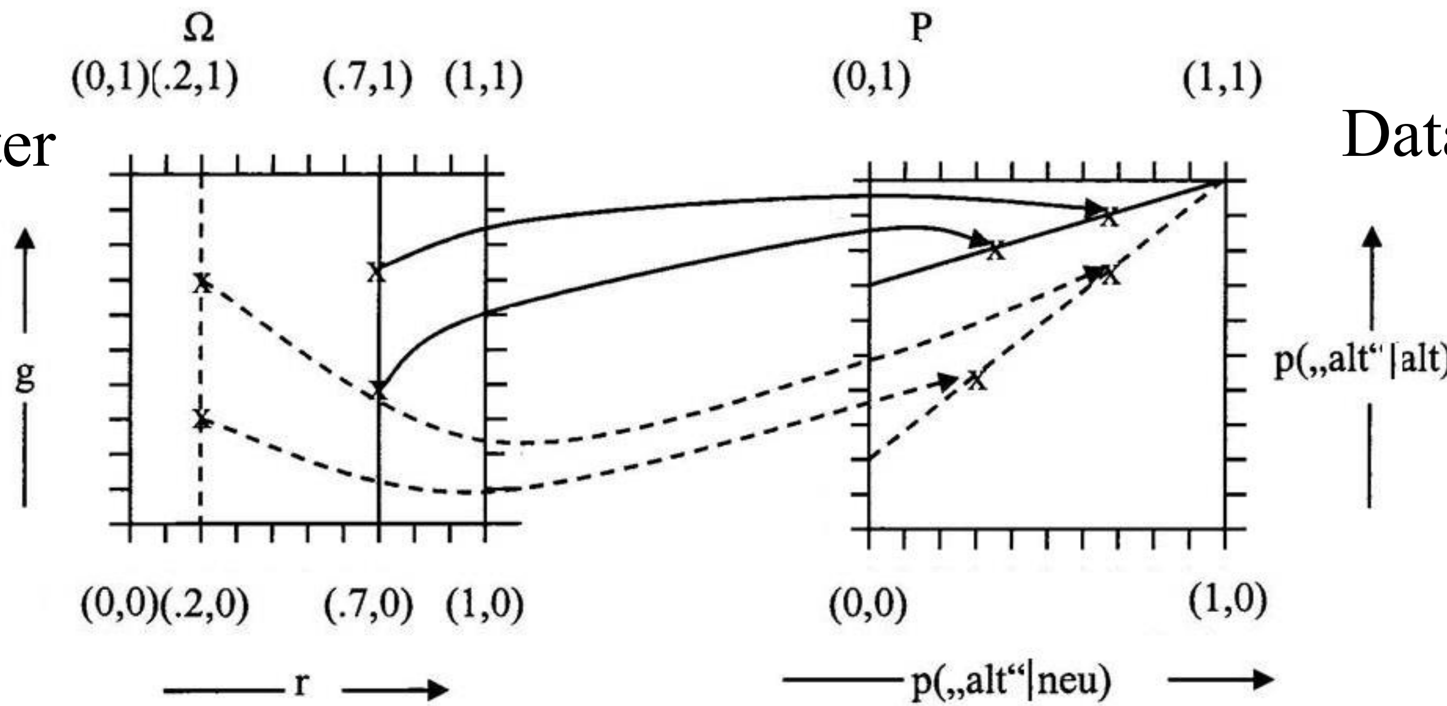
Example 1: Nonidentifiability



- One-High-Threshold-Modell limited to old items:
  p(“old” | old item) = $r + (1\text{-}r) \cdot g$
- Model is *not* identified.

Example 2: Identifiability

Parameter space



Data space

- One-high-Threshold-Model (old and new items)
  - p("alt" | old item) $= r + (1\text{-}r) \cdot g$
  - p(„alt“ | new item) $=$ g
- Model *is* globally identified.

Identifiability: Two Important Theorems

- „*Observable branches*“:

  A model is always globally identified if each of its branches terminates in a new empirical category (Hu & Batchelder, 1994).

- „*No more parameters than degrees of freedom in the data*“:

  A necessary but not sufficient condition of identifiability for the number of parameters *S* is:

$$S \leq \sum_{k=1}^{K} (J_k - 1)$$

Identifiability: Jacobian Matrix

- Jacobian: Matrix of the first partial derivatives of all model equations with respect to all parameters θ_s $(s = 1, \ldots S)$
- r: maximum rank of the Jacobian across Ω
- If $r < S$, then the model is neither locally nor globally identified.
- If $r = S$, then the model is locally identified (but not necessarily globally).

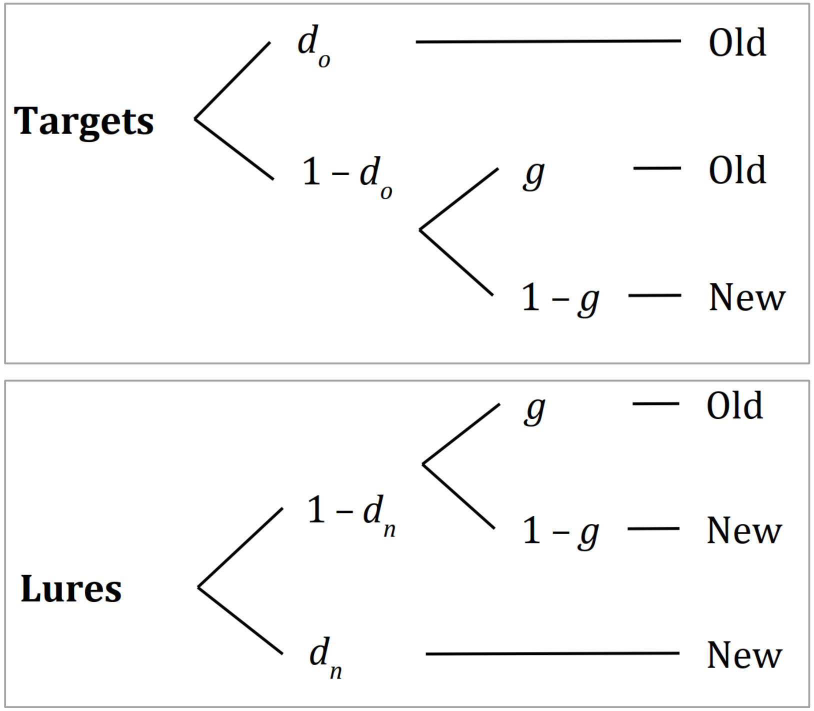
Remedies for Nonidentifiable Models

- Less parameters → Parameter constraints
  – Parameter fixations ($\theta_s = c$, with c = constant)
  – Equality constraints ($\theta_s = \theta_t$)

- Increase the number of empirical categories
  – Additional conditions with no (or few) additional parameters
  – Selective manipulations of parameters

1.5) Identifiability : Example

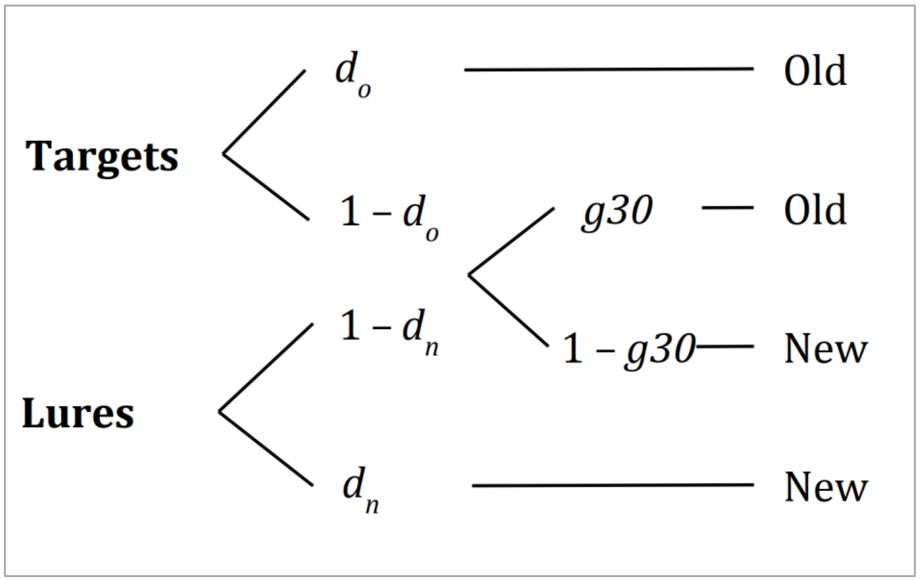
- 2-High Threshold Model
  - Parameters: $S = 3$
  - Free categories: $df = 2$
  - → not identifiable!



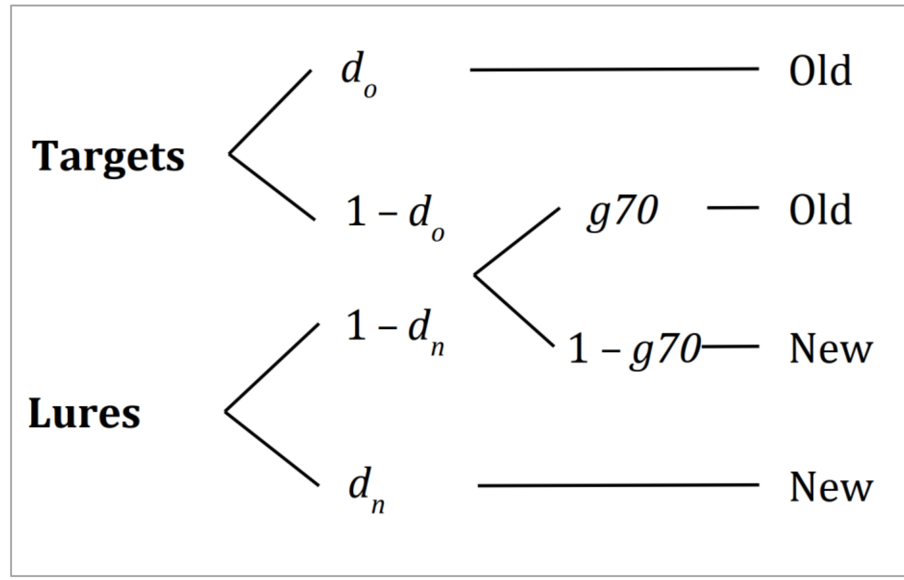
- Solutions
  - Assume $d_o = d_n$
  - Base rate manipulation of response bias g

Increasing the number of empirical categories

(A) Base rate: 30% targets



(B) Base rate: 70% targets



→ Two additional degrees of freedom ($df = 4$)
→ But only one more free parameter ($S = 4$)

1.6) Parameter Estimation

- Given the data $n_1, \ldots, n_J$, what is the „best" vector of parameter values $\boldsymbol{\theta} = (\theta_1, ..., \theta_s, ..., \theta_S)$?
  → Find $\boldsymbol{\theta}$ that mimimizes the distance between observed and expected category frequencies!

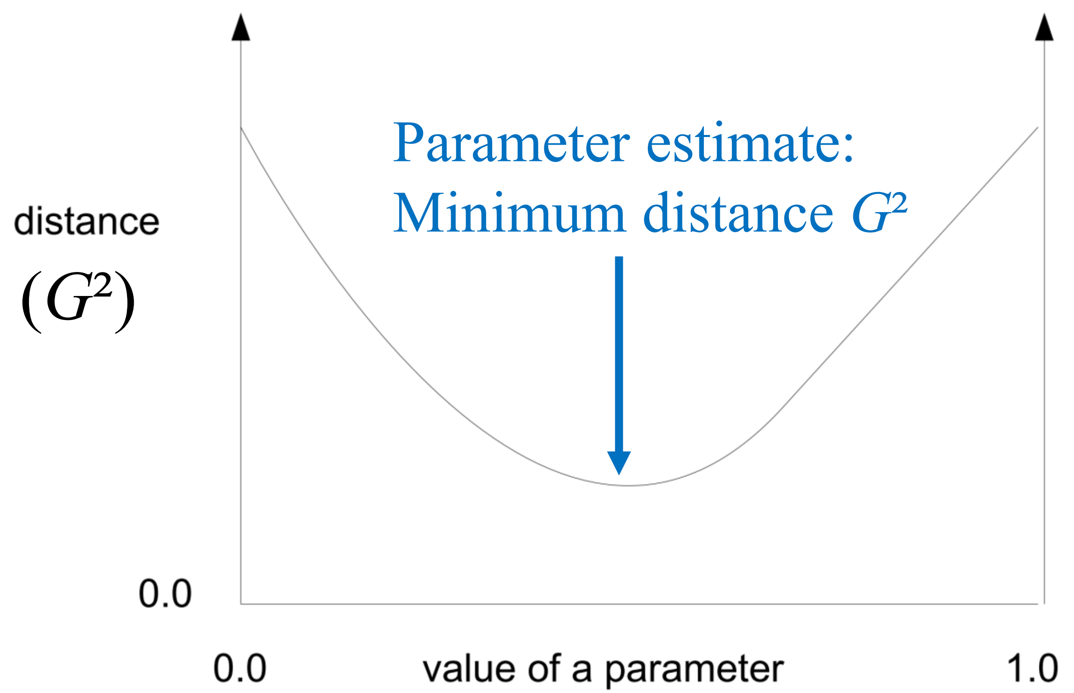
- Distance measure: The likelihood ratio statistic G^2

$$G^2(\boldsymbol{\theta}) = 2\sum_{j=1}^{J} n_j \ln\left(\frac{n_j}{N \cdot p_j(\boldsymbol{\theta})}\right)$$

- Note: Minimizing G^2 is equivalent to maximizing the likelihood of the data given the parameters.

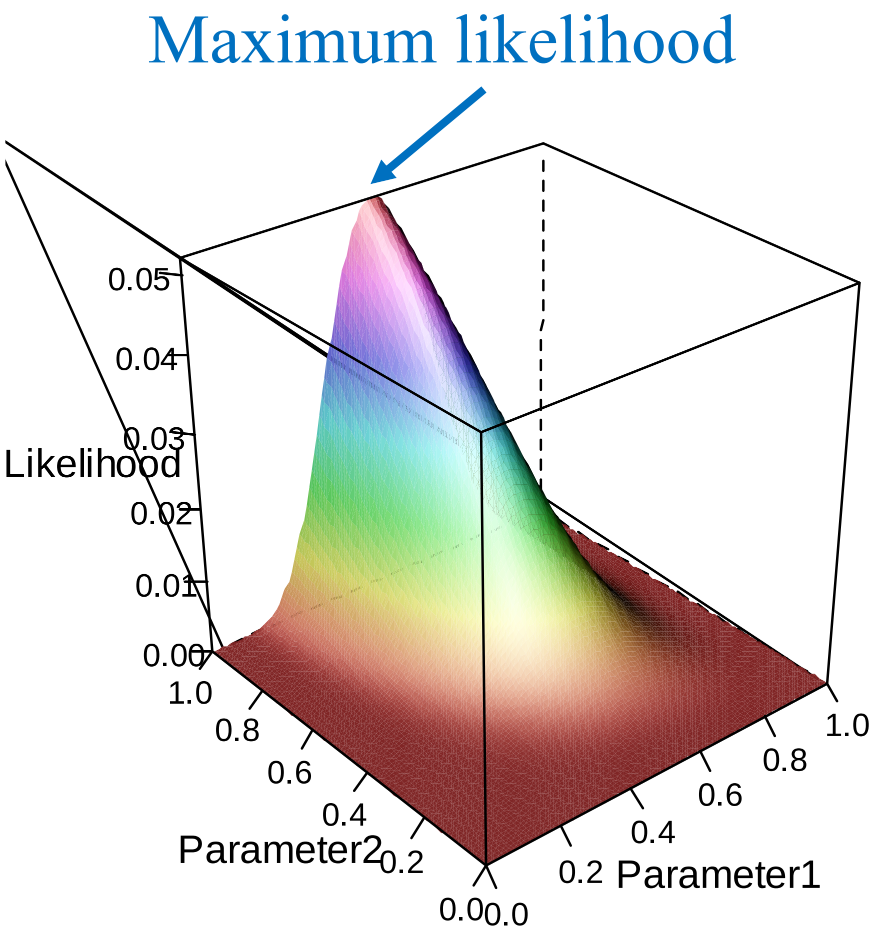
1.6) Parameter Estimation

- Minimization of the distance G^2 for an MPT model with $S = 1$ parameter:



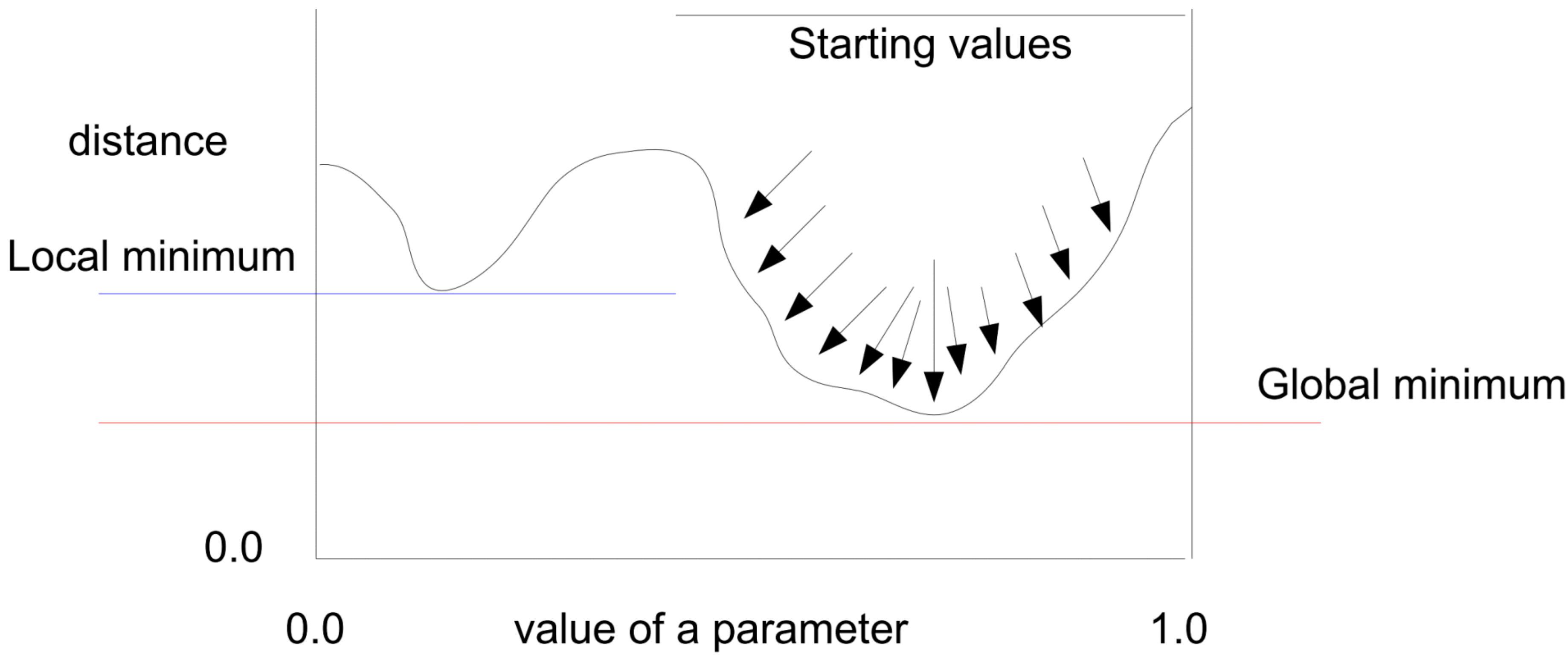
1.6) Parameter Estimation

- Parameter estimation for an MPT model with $S = 2$ parameters:

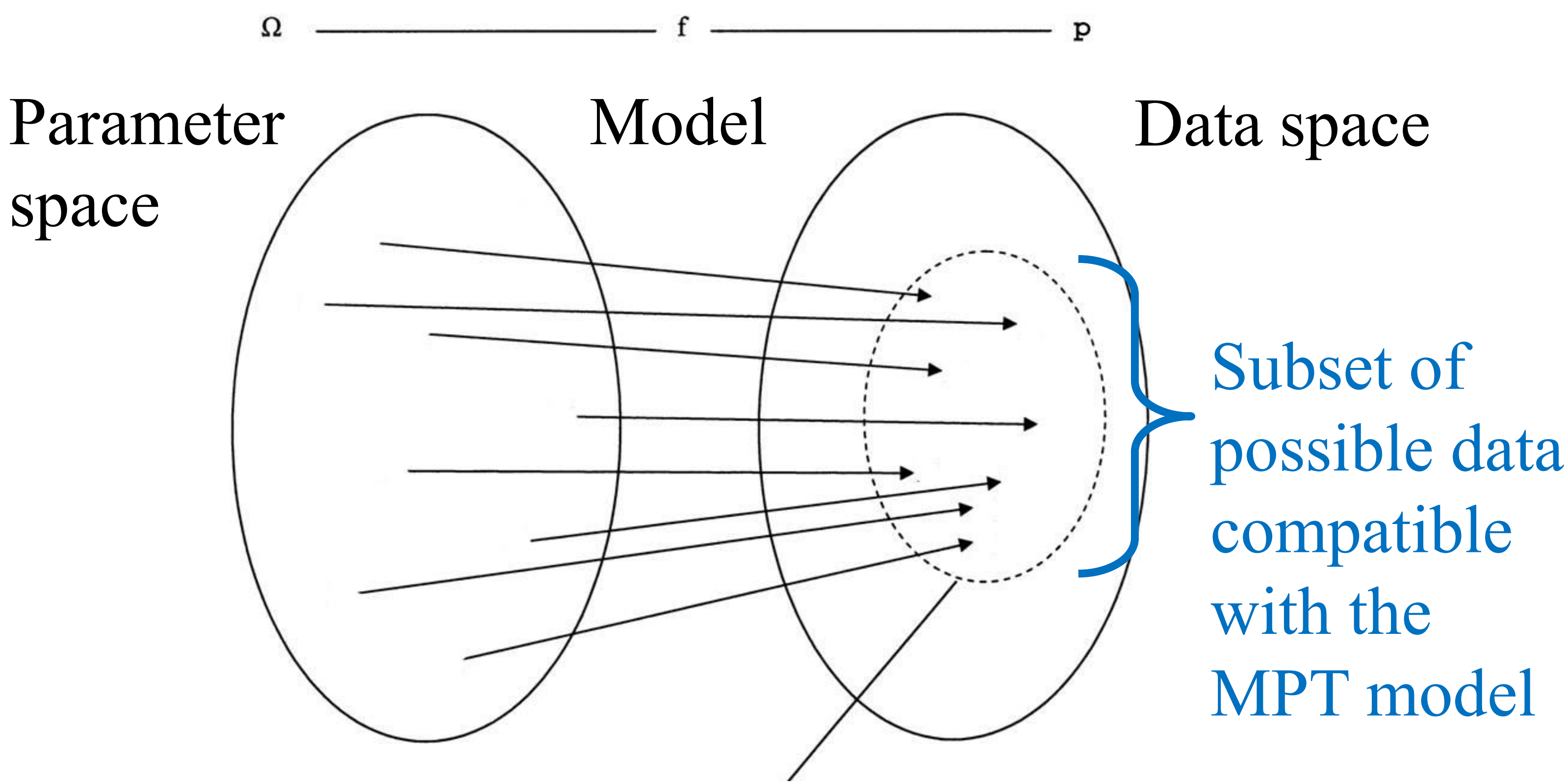


Parameter Estimation: Local Minima

- Solution to cope with local minima:
- Fit model multiple times with random starting values



1.7) Model Assessment



1.7) Model Assessment

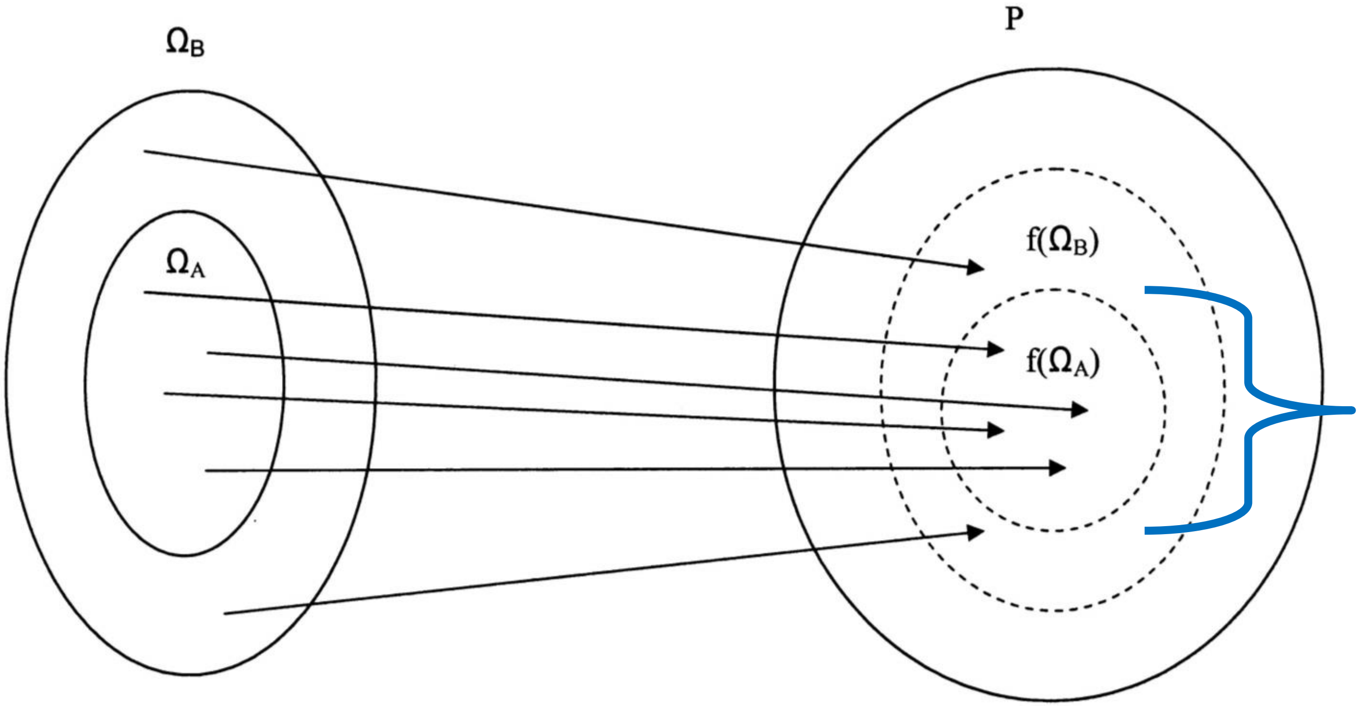
- How to test whether a model fits the data?
  - Hypothesis: the data are generated by the model
  - H_0: $\pi \in f(\Omega)$
- In this case, G^2 is χ^2-distributed with degrees of freedom:

$$df = \sum_{k=1}^{K} (J_k - 1) - S$$

1.7) Model Comparisons: Nested Models

Hierarchical model families:

- Model M_A is a special case (a nested model) of M_B
- e.g., M_A is obtained from M_B via parameter restrictions



Subset of possible data compatible with both MPT models

Model comparisons in hierarchical model families

- If model M_A is nested in M_B then
  - G^2_A is χ^2-distributed with df_A
  - G^2_B is χ^2-distributed with df_B
  - $\Delta G^2_{A\text{-}B} = G^2_A - G^2_B$ is χ^2-distributed $df_{A\text{-}B} = df_A - df_B$

- Hence, we can use $\Delta G^2_{A\text{-}B}$ to compare nested models using χ^2-tests

- To compare non-nested models, we can use information-theoretic measures (AIC, BIC)